DOCUMENT RESUME

ED 223 717 TM 820 864

AUTHOR McLarty, Joyce
TITLE Technical Considerations in Quantifying Severe Discrepancies for the Identification of Learning Disabilities.
INSTITUTION Riverside Unified School District, Calif.
PUB DATE 11 Nov 82
NOTE 28p.
PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; Correlation; *Evaluation Methods; *Handicap Identification; *Learning Disabilities; Mathematical Models; *Measurement Techniques; *Severe Disabilities; Special Education; State Standards; Statistical Analysis; Student Characteristics; Test Construction; Test Reliability
IDENTIFIERS California

ABSTRACT

Educators designing procedures for identifying students as learning disabled are faced with serious technical difficulties. A student with a learning disability is one who "has a severe discrepancey between achievement and intellectual ability" in one or more designated areas. The methods of quantifying achievement, intellectual ability, and their difference together with the nature and degree of difference specified as "severe" determine, in large part, the number and characteristics of students identified as learning disabled. Three types of approaches are currently used for the quantification of severe discrepancies: absolute difference methods, expected achievement methods, and proportional difference methods. Each approach may be utilized with a variety of statistical models. The most complex models adjust the quantification and standard for many characteristics of the student and of the tests. The three approaches are reviewed with reference to some of the commonly used statistical techniques. The effects of test reliability, regression to the mean, test intercorrelations, multiple comparisons, Type I and Type II error rates, and the choice of student and test characteristics for which adjustments in the characteristics of students identified as learning disabled are considered. The advantages and disadvantages of the model proposed by the state of California are discussed. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223717

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. McLarty

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

RIVERSIDE UNIFIED SCHOOL DISTRICT
Riverside, California

DEPARTMENT OF RESEARCH AND EVALUATION
November 11, 1982

TECHNICAL CONSIDERATIONS IN QUANTIFYING SEVERE DISCREPANCIES FOR THE IDENTIFICATION OF LEARNING DISABILITIES

Joyce McLarty, Ph.D.
Research Associate

TM 820 864

Educators who design procedures for identifying students as learning disabled encounter serious technical problems. By law, a student with a learning disability has "a severe discrepancy between achievement and intellectual ability" in one or more designated areas. This discrepancy, linked to specific deficits in psychological processing, must not be the result of other types of handicapping conditions, lack of effort or instruction, or non-standard language or cultural background. Procedures and standards for quantifying a "severe discrepancy" therefore identify students who may, if alternative causes of the discrepancy can be ruled out, be classified as learning disabled.

Conceptually the appropriate procedure is clear: (1) measure the students' intellectual ability; (2) measure the students' achievement level; (3) calculate the difference between the students' ability and achievement, and (4) compare this difference to a standard representing "severity." Each of these steps presents technical difficulties. Measurement of intellectual ability and of achievement in Steps 1 and 2 require attention to the reliability and validity of the tests used. Since students referred for suspected learning disabilities can be expected to have low achievement scores (and, to the extent that ability and achievement are correlated, low ability scores as well), special attention

to technical concerns such as test "floors" and the effect of truncated distributions on test statistics may also be required. Students are referred for suspected learning disabilities because they perform more poorly in some specific area than in other areas. Many of these students also have large subtest discrepancies on ability measures. To which ability score (full scale, verbal, performance) should the achievement scores be compared?

In Step 3, the difference between ability and achievement must be calculated. On which common scale and with what assumptions about underlying distributions should the difference score be examined? What is the effect on the differences of the tests not having been co-normed? The distribution of difference scores is affected by the reliability (or the lack thereof) of both tests, and also by the correlation between them. If ability provides a theoretical "ceiling" for achievement, the distribution of achievement conditional on ability will be not only heteroscedastic (variance increasing with ability) but negatively shewed as well. What effect does this have on the numbers and characteristics of identified students?

In Step 4, the difference scores must be compared to a standard representing "severity." Should the standard be norm-referenced or criterion-referenced? How will it be selected? How many students will it identify? What will the characteristics of identified students be? What proportions of Type I and Type II errors can be expected? How can these errors be minimized?

In addition to considering the appropriate identification procedure for individual students, it is important to consider the referral and assessment systems now in place. Students referred for suspected learning disabilities are generally administered an individual ability test by a school psychologist or other professional. The most commonly used tests are the WISC-R and the Stanford Binet, but at least a half dozen other tests are in regular use. An individual achievement test is often administered, but possibly at a different time than the ability test and by another tester (for example the classroom teacher). Again there are several commonly used tests (WRAT, PIAT, Woodcock-Johnson), but many other tests are in regular use. Some of the achievement tests are too new for there to be much data on them. For other tests, technical data may be more than ten years old. Criterion-referenced achievement tests may lack norming information. Research which relates ability to achievement is generally based on referred populations with restricted score ranges. It is available for only certain pairs of tests and the research designs may leave doubts about the accuracy of the information reported.

Despite the numerous and complex technical difficulties, a number of different methods for quantifying severe discrepancies for the identification of learning disabilities have been designed and are in use. They vary considerably in technical sophistication. The simplest methods attend to few of the technical aspects of the data and the decision system; the most complex attempt to adjust for many. The methods may be categorized along several dimensions, but tend to cluster into three general categories: absolute difference methods, expected achievement methods, and proportional difference methods.

Absolute Difference Methods

Absolute difference methods ignore the relationship between ability and achievement measures. Characteristics of each measure are considered, but separately. A simple absolute difference approach is the years-below-grade-level standard (Erickson, 1975). A student whose achievement is a given number of grade-equivalent units below grade placement is identified. The number of "years-below" selected as a standard may be held constant, or it may be increased with grade level to compensate for normal increases in the variance of achievement scores. Since no adjustment is made for ability, this method does not discriminate between slow learners and the learning disabled.

A more sophisticated absolute difference method is the standard or z-score approach (Elliott, 1981). Ability and achievement scores are converted to a standard scale and their difference is compared to a selected standard on the same scale (for example, a z-score of 1.5). With this method, adjustments are made for the characteristics of each test but not for the relationship between them. The conversion procedure automatically compensates for increases in score variation within age/grade. No correction is made for changes in the variation of achievement scores conditional on increased ability, so this method can be expected to identify more high than low ability students.

Expected Achievement Methods

Expected achievement methods quantify expected achievement based on selected characteristics of the test and individual and on an explicit model of expectation. They then compare the difference between expected

and actual achievement to a selected standard. A simple expected achievement approach is that proposed by Harris (1970): the expected grade equivalent is equal to the child's mental age minus five years. Children whose actual achievement is below their expected achievement are identified. More complex, but still intuitive formulae include the following:

(1) Expected Grade Equivalent = Years in School x $\frac{IQ}{100}$ +1.0

Bond and Tinker (1967)

(2) Severe Discrepancy = CA ($\frac{IQ}{300}$ + .17) = 2.5

Cited in Shepard (1980)

(3) Expectancy Age = ($\frac{2MA + CA}{3}$)

(4) Expected Grade Equivalent = $\frac{CA\ Months}{100}$ x IQ - 12 - 5.3

State of Florida

Where CA = chronological age

MA = mental age

IQ = intelligence quotient

The typical effect of such formulae are standards which are inconsistent with respect to both age/grade and ability. Table 1 illustrates this effect. In this case, the standard is set as 75 percent of expected achievement using the State of Florida formula which is

expressed in grade equivalent units. To generate this table, the grade equivalent standard was computed for the selected age/IQ combinations then expressed in scaled score units. Since the WISC-R and WRAT have the same scaled score metric (mean = 100, standard deviation = 15), it was possible to express the difference between the IQ and the maximum qualifying achievement score by direct subtraction. The resulting difference varies by both age and IQ in an interactive fashion. The qualifying difference is smaller for the oldest students (grade 11 and up) and the youngest (grade 1 and below) and for those with lower IQs.

Because the formulas are experientially rather than empirically based, the patterns of standards they produce vary considerably.

TABLE 1

MINIMUM SCALED SCORE DISCREPANCIES FOR WISC-R AND WRAT READING USING EXPECTED GRADE EQUIVALENT FORMULA[1]

	IQ					Average Diff.
	80	90	100	110	120	
Grade: 1		Age: 6				
Maximum Grade Rating	P.6	K.1	K.5	1.0	1.4	
Scaled Score Equivalent	72	81	86	95	103	
Difference	8	9	14	15	17	12.6
Grade: 3		Age: 8				
Maximum Grade Rating	K.8	1.4	2.0	2.6	3.2	
Scaled Score Equivalent	60	68	75	87	99	
Difference	20	22	25	23	21	22.2
Grade: 5		Age: 10				
Maximum Grade Rating	2.0	2.8	3.5	4.3	5.0	
Scaled Score Equivalent	62	78	87	94.5	99	
Difference	22	12	13	15.5	21	16.7
Grade: 7		Age: 12				
Maximum Grade Rating	3.2	4.1	5.0	5.9	6.8	
Scaled Score Equivalent	70	78	84	91.5	100	
Difference	10	12	16	18.5	20	15.3
Grade: 9		Age: 14				
Maximum Grade Rating	4.4	5.5	6.5	7.6	8.6	
Scaled Score Equivalent	72.5	81.5	89.5	97	106	
Difference	7.5	8.5	10.5	13	14	10.7
Grade: 11		Age: 16				
Maximum Grade Rating	5.6	6.8	8.0	9.2	10.4	
Scaled Score Equivalent	80	89	98	107	117	
Difference	0	1	2	3	3	1.8
Average Difference	11.25	10.75	13.4	14.7	16	

[1]Standard is 75 percent of Expected Grade Equivalent = $\left| \frac{\text{CA months}}{100} \times \text{IQ} \div 12 \right| - 5.3$

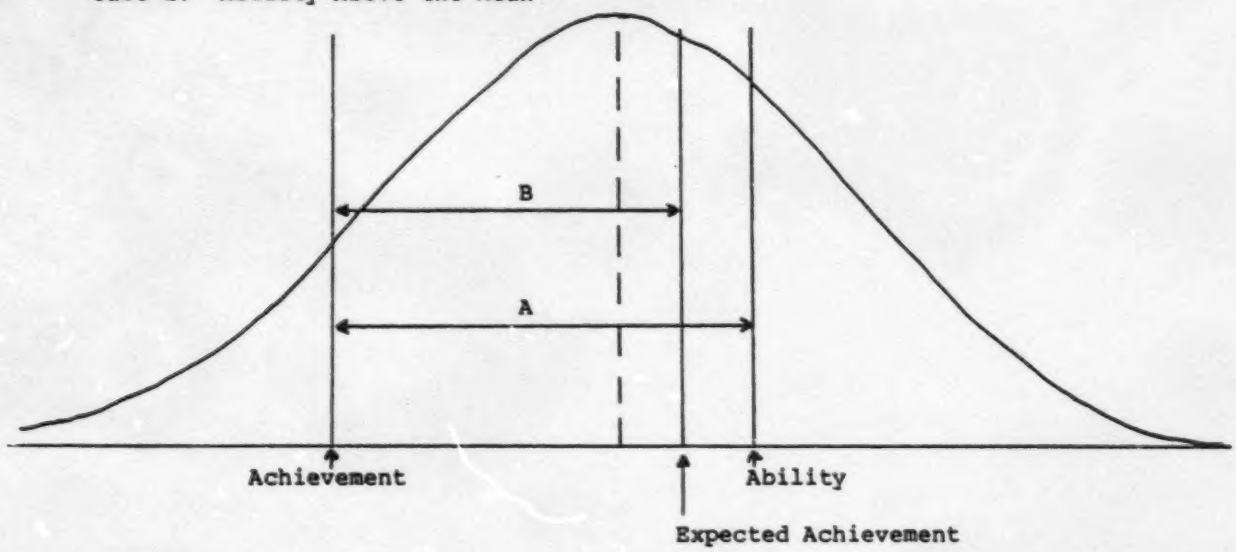
A more sophisticated version of the expected achievement approach is the regression method (Shepard, 1980) which calculates a student's expected achievement using an empirically derived prediction equation based on previous statistical studies. The difference between expected and actual achievement is then compared to a selected standard. This method adjusts for the relationship between ability and achievement measures as well as for those student characteristics included in the regression prediction equation(s).

It also tends to identify a greater proportion of low ability students than do other methods. This is because the expected achievement score to which the actual achievement is compared is regressed toward the mean, increasing the difference score if ability is below the mean and decreasing it if ability is above the mean (see Figure 1). In the first case since ability is above the mean, expected achievement is below it. This reduces the discrepancy from length "A" (ability-achievement difference) to length "B" (achievement-expected achievement difference), making it less likely that the higher-ability child will be identified. In the second case, with ability below the mean, expected achievement is above ability, which increases the observed difference from length "A" (ability-achievement difference) to length "B" (ability-expected achievement difference), making it more likely that the lower-ability child will be identified.

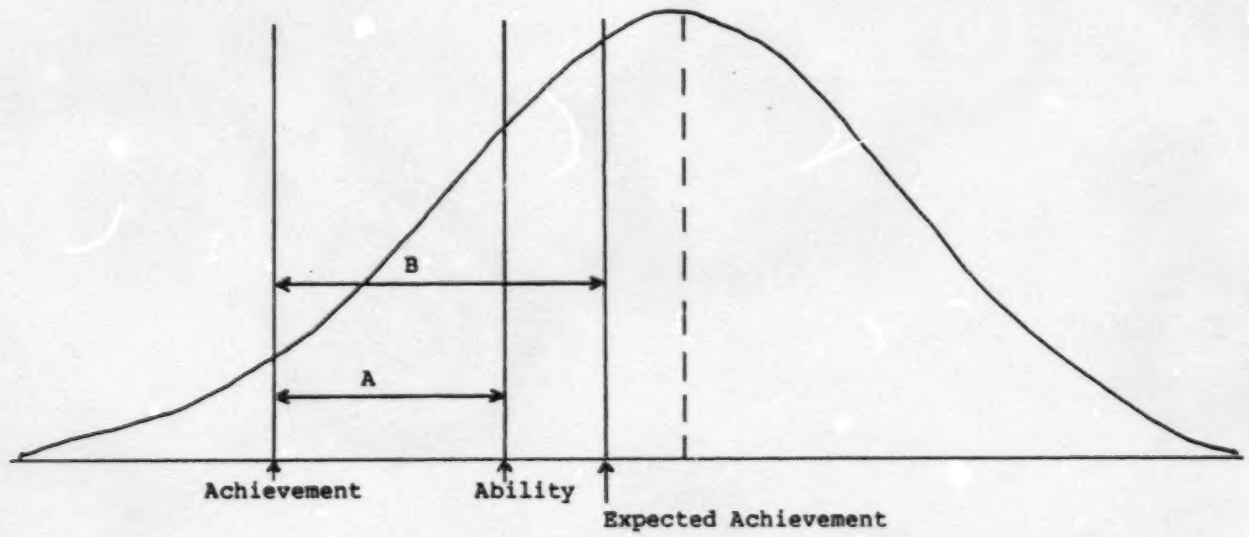
Regression methods predict achievement above ability for the low ability child. In addition to being conceptually awkward, this approach may be theoretically inappropriate. This is because the regression-

Figure 1

Case 1: Ability Above the Mean



Case 2: Ability Below the Mean



prediction assumes that the students were selected on the basis of low ability (predictor) scores when in fact they were probably referred because of a history of poor achievement.

Regression methodology does not adjust for changes in score variation associated with ability (which tend to result in more high than low ability students being identified). The net effect of regression methodology on the proportions of students selected at each ability level therefore depends on the relative strength of the relationships between (1) the ability and achievement measures and (2) the ability measure and the variance in achievement scores.

Proportional Difference Methods

Proportional difference methods consider the distribution of differences, rather than the distributions of test scores (actual or expected). Such methods focus on the variation (standard deviation) of the difference scores or on their error variation (standard error). In either case, the simplest methods compare the actual difference, expressed on any convenient scale, to a selected proportion of the variation (actual or error). A standard which refers to the error variation scale identifies students with statistically significant differences between ability and achievement. A standard referred to the standard deviation scale identifies students with unusually large differences between ability and achievement in comparison to the normative group. Usually the standard is multiplied by a constant which expresses the severity criterion. This multiplier relates closely to the proportion of students identified for a standard deviation scale and to the probability of making a Type I error for a standard error scale.

A more sophisticated proportional difference method would adjust the standard conditional on the level of the ability measure. Since the variance of difference scores increases with ability level, the standard deviation/standard error could be calculated separately for each level (the proportion used as the multiplier in arriving at the standard would remain constant).

Shepard (1980) recommends a regression-based expected achievement method combined with a standard in which the lowest 10 percent of the students at each ability level are identified. This combines a linear prediction method with a correction for changes in the variation of difference scores with increased ability levels. It also requires that the same percentage of students at each ability level be identified.

State of California Method

The State of California is proposing a proportional difference method based on the standard deviation of the difference scores. This method adjusts for the correlation between the measures of ability and achievement and therefore, indirectly, for their reliabilities. The recommended standard is 1.5 standard deviations of the difference distribution. For any single pair of tests, this method can be expected to identify the 6.68 percent of the normal population with the largest ability-achievement differences. This percentage increases as multiple pairs of tests (ability with reading, math, language, etc.) are considered. Because no adjustment for changes in score variation with ability is made, this method will identify more students with above than below average ability.

Adjustments for Student and Test Characteristics

Discrepancy models vary considerably in the ways in which they adjust for characteristics of the student and test. Student characteristics are most commonly handled by referring the student's score to an appropriate norms table. This compensates for differences in age/grade (there is some question as to whether the age or grade level table is more appropriate, especially for students who have been retained). Expected achievement methods also provide adjustments for the students' ability level (e.g., by utilizing ability information in calculating an expected achievement level).

Regression approaches may provide adjustments for the student's sex, ethnic group or other background information as well as for ability levels. The selection of variables to be included in the regression equations is based on theoretical and statistical considerations. Variables which are not predictive of achievement need not be included in the equations, even if they are theoretically relevant. Those which are not theoretically relevant should not be included, even if they are predictive. The inclusion of a characteristic positively associated with achievement (e.g. high SES) will result in a greater proportion of students with that characteristic being identified. Conversely, the inclusion of a characteristic negatively associated with achievement (e.g. minority language status) will reduce the proportion of students with that characteristic identified.

Adjustment for test characteristics begins with converson of the scores to a common scale. This compensates for differences in test means and variances. Expected achievement and proportional difference methods

also adjust for the relationship between ability and achievement on an intuitive or statistical basis. Test reliability is usually considered in standard-setting either by utilizing a standard based on statistically significant differences (standard error of the difference scores) or by placing a confidence interval around the cutoff score. This confidence interval provides a band within which it is not possible to determine whether the discrepancy criteria has been met because of the unreliability of the tests. Where test reliability is poor or tests are highly correlated, this band may be very large.

The correlation between the ability and achievement measures is an important determinant of the likelihood of large observed differences between ability and achievement scores. If the tests are not highly correlated, large differences will be common; if the tests are closely related, large differences will be rare. Figure 2 illustrates the effect of the correlation between the ability and achievement test on the standard deviation of the difference distribution. The formula (in z-score units) is:

$$SD_{dif} = \sqrt{2\ (1 - r_{xy})}$$

where r_{xy} is the correlation between the two tests. When the correlation between the tests is .50, the standard deviation of the differences is equal to the standard deviation of their common z-score scale, that is, 1.0. As the correlation between the tests increases, the standard deviation of the difference decreases in an accelerating fashion and approaches zero. Methods which do not adjust for differences in the correlation between different pairs of ability-achievement tests will tend to identify many students who were administered tests with low correlations and few students who were administered highly correlated tests.

The correlation between the two tests also plays an important role in determination of standard errors. The standard error of measurement and the standard error of estimate are statistics which reflect the accuracy of a measurement or a regression estimate, respectively. Identification strategies which focus on a standard error term have the unique advantage of permitting estimates of the proportions of classification errors. The proportion of Type I errors (observation of a significant discrepancy when no disability exists) is determined directly by the standard (one standard error will incorrectly identify about 16 percent of the population, two standard errors will identify two percent, and three standard errors will identify .13 percent).

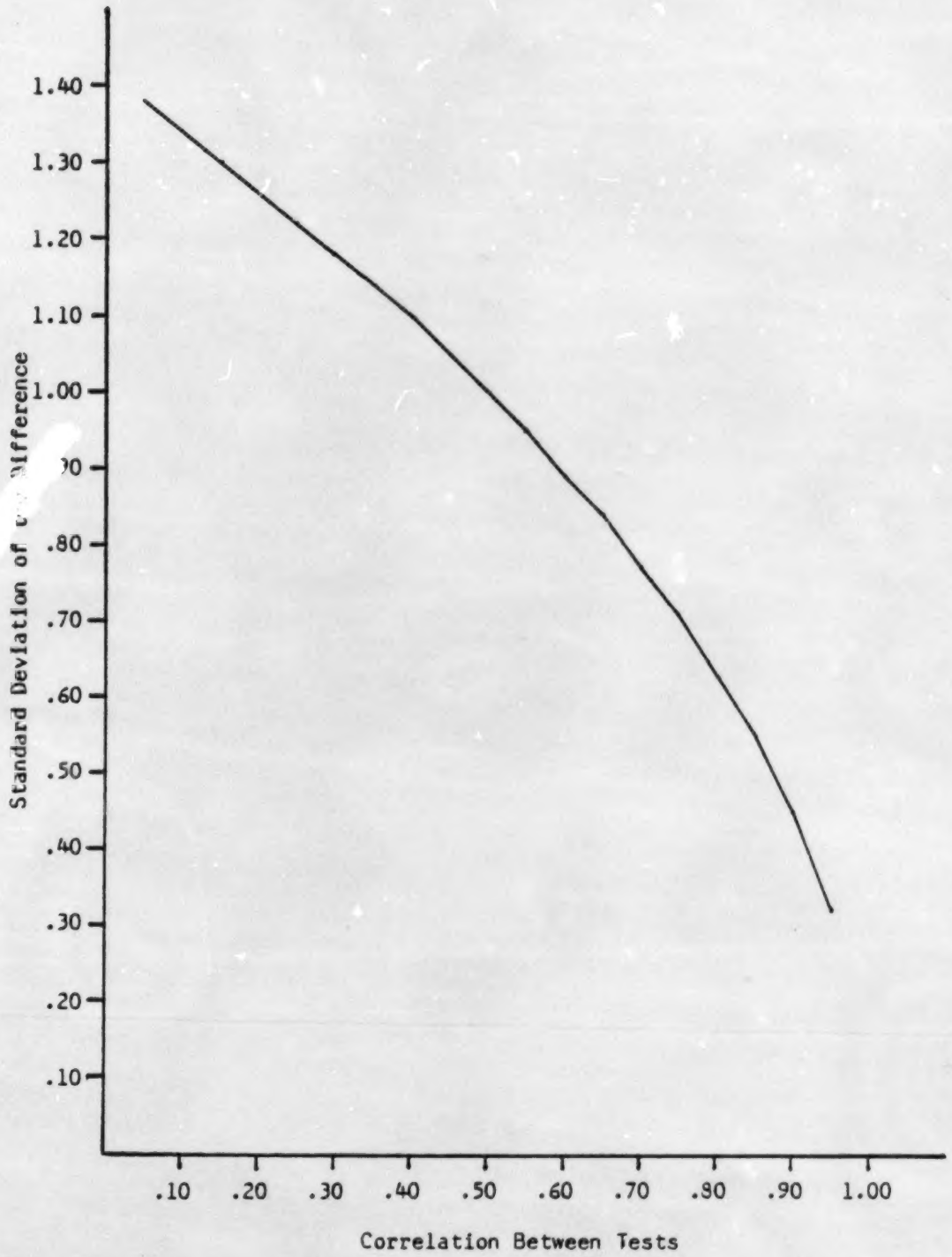
The prevalance of Type II errors (failure to detect a disability-related discrepancy) depends on the severity of the disability. A severe disability is easy to detect, a mild one may be very difficult to detect. Since the size of the difference to be detected is unknown, the usual method of investigating Type II error rates is to identify the size of difference (effect) which it would be important to detect. This effect is statistically related to the sample size, the metric in which the scores are expressed (mean and standard deviation), and the Type I error rate. The formula is:

$$n = \frac{(z_{\alpha} - z_{\beta})^2}{(\mu - \mu_0)^2 / \sigma^2}$$

Kirk, p. 230

Figure 2

THE EFFECT OF INTER-TEST CORRELATIONS ON THE STANDARD DEVIATION OF DIFFERENCE SCORES



where:

n = sample size

z_a = Type I error rate

z_B = Type II error rate

$u - u_o$ = size of difference, in original scale units

o^2 = variance of original scale

Assuming a standard (z-score) scale with a mean of zero and variance of one, and a sample size of one (the decision is to be made for each student), this formula reduces to:

$$z_a = u + z_B$$

that is, the Type I (alpha) level is equal to the sum of the Type II (beta) error level and the size of the effect (difference) each expressed in standard score units. For a desired power level of .80 (80 percent probability of detection), it is possible to detect a difference of about -2.5 with alpha equal to .05, or a difference of -3.2 with alpha equal to .01 (one-tailed tests).

The effect of the sample size of one is to make it impossible to identify any but the most severe discrepancies without also identifying large proportions of normal students. This situation is exacerbated by the low incidence of learning disabilities in the general population (Shepard, 1980). Given the same proportions of misidentification, the differences in incidence assure that more normal students will be identified as having severe discrepancies than learning disabled students identified as normal.

Figures 3 and 4 illustrate the effect of the correlation coefficient on the standard error of measurement and the standard error of estimate, respectively. The standard error of measurement is affected by the correlation between the two tests and their individual reliabilities. The correlation between the tests cannot conceptually exceed the lower of the two reliabilities. Figure 3 plots the standard error terms for the average reliability of the two tests, such that the correlation doesn't exceed this average.

This provides a lower bound for the standard error of measurement (under the condition that the two tests are equally reliable). If one test is more reliable than the other, the table underestimates the size of the error. The formula is:

$$SEM_{dif} = \sqrt{1 - r_{dd}}$$

$$\text{where } r_{dd} = \frac{\frac{1}{2}(r_{xx} + r_{yy}) - r_{xy}}{1 - r_{xy}}$$

(after McNemar, p. 175, 10.24)

where r_{xx} and r_{yy} are the test reliabilities and $\frac{1}{2}(r_{xx} + r_{yy})$ is their average; r_{xy} is the correlation between the tests; and r_{dd} is termed the reliability of the difference.

The standard error of measurement approach identifies students for whom the ability/achievement discrepancy is too large to have occurred by chance. The standard error of estimate identifies students for whom the expected-achievement/achievement difference is too large to have occurred by chance, where expected achievement has been estimated using a regression model.

Figure 3

THE EFFECT OF MEAN RELIABILITY AND INTER-TEST CORRELATIONS ON THE STANDARD ERROR OF MEASUREMENT OF THE DIFFERENCE SCORES

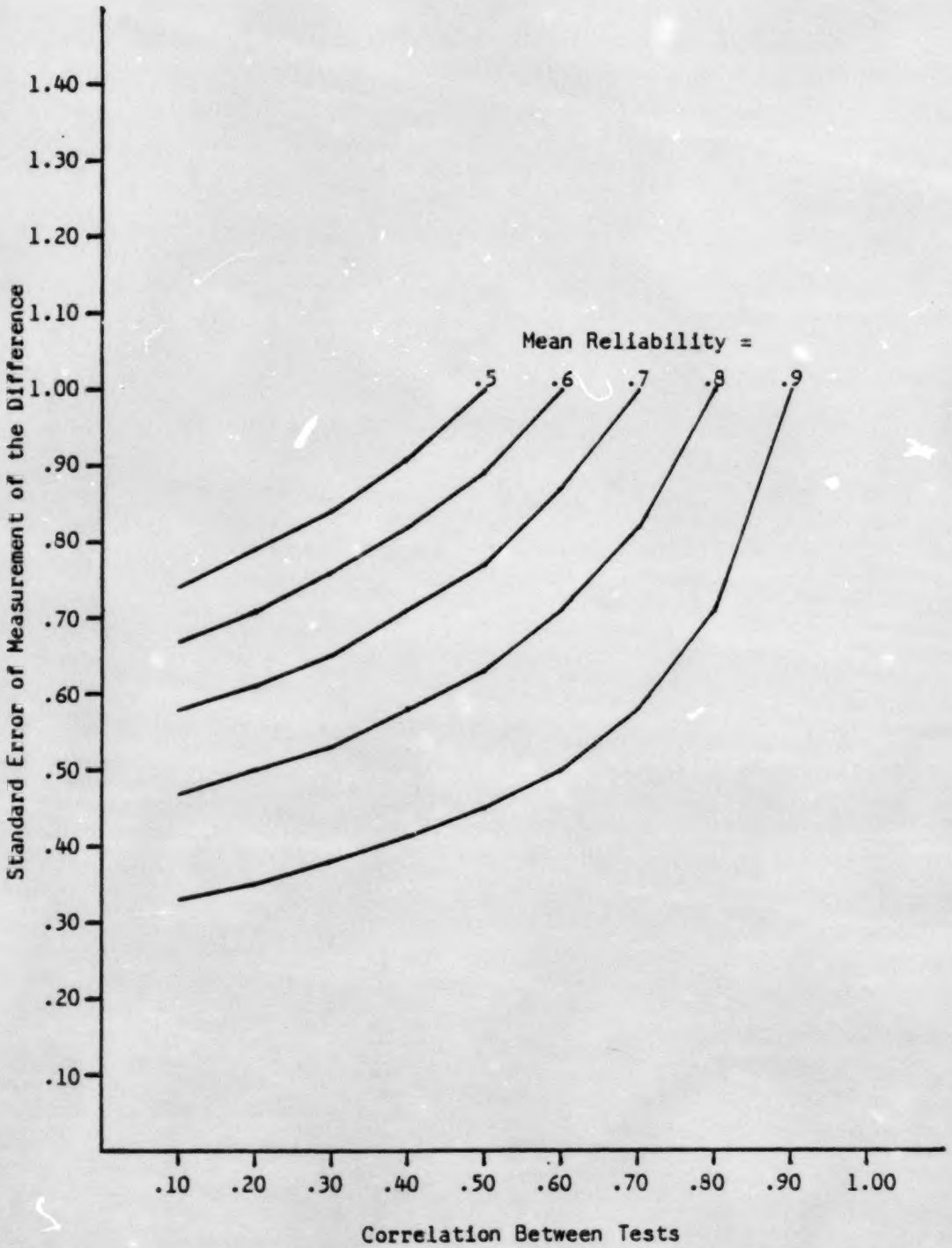
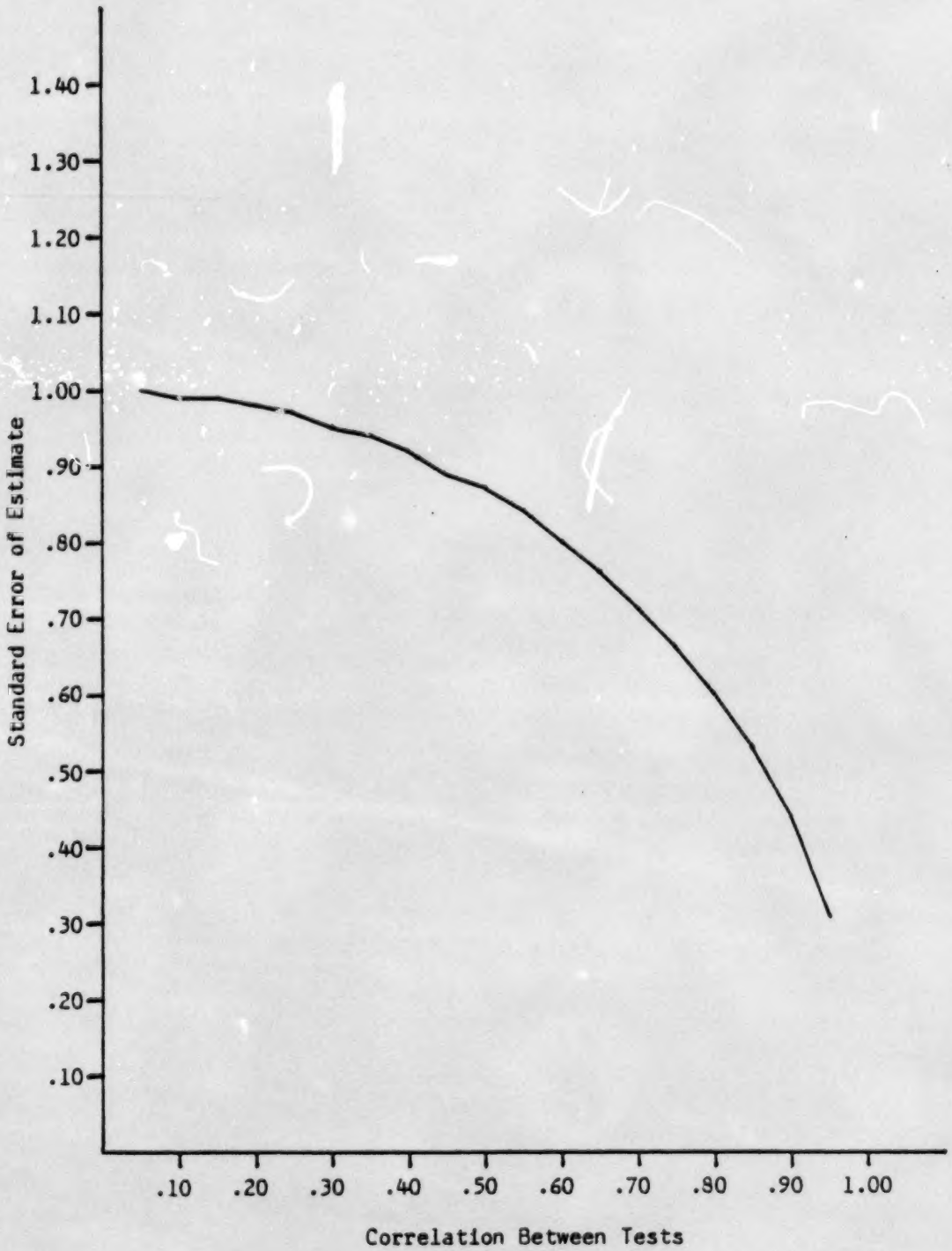


Figure 4

THE EFFECT OF INTER-TEST CORRELATIONS
ON THE STANDARD ERROR OF ESTIMATE



The formula on the z-score scale involves only the correlation between the two tests:

$$SE_{est} = \sqrt{1 - r^2_{xy}}$$

(after McNemar, p. 139, 9.6)

Comparison of the standard deviation of the difference and the standard error of estimate indicates that as the correlation between the tests increases, greater and greater proportions of the difference distribution are within the error of estimate distribution until as the correlation approaches 1.0 the two distributions are synonymous.

The standard error of measurement functions somewhat differently because it depends on both the correlation between the tests and their reliabilities. The standard error of measurement decreases as test reliability increases; but for a given reliability level, the standard error of measurement increases with the correlation between the tests. When the correlation is equal to test reliability, the standard error of measurement is equal to 1.0.

These three distribution indices (standard deviation of the difference, standard of error of estimate and standard error of measurement) have in common that their values change most rapidly in the upper ranges of the correlation coefficients. The relationship between the standard error of measurement and the other two statistics depends on the reliabilities and inter-test correlation involved.

Ability/Achievement Correlations

Determining the reliabilities of and especially the correlations between commonly used ability and achievement tests is itself a technical problem. Test manuals and the research literature report not one but many reliabilities and correlations. These vary primarily with respect to subject characteristics (age/grade, ethnic group, sex, reason for referral). They may also vary because of specific research design considerations (test sentization, regression effects), because of the age of the study, or because of the type of statistic calculated (e.g., internal consistency versus test-retest reliability). For some tests little data is available. For many pairs of tests, the variation in reported correlations appears too large to have occurred by chance. Examination of the studies reporting these correlations does not suggest systematic differences relating to the degree of range restriction or the quality of the study.

The overall pattern is one of relatively high reliabilities (.7 to .8) coupled with moderate correlations (.3 to .6). In this range, the standard error of measurement is smaller than the standard error of estimate, which is smaller than the standard deviation of the difference. All three statistics converge as the correlation between the tests increases, then diverge again as the standard error of measurement exceeds the standard deviation of the difference (see Figure 5). When the correlation between the tests exceeds .5, all three statistics are smaller than a z-score of 1.0 (one standard deviation). When the correlation is below .5 only the standard deviation of the difference exceeds 1.0.

Figure 5

RELATIONSHIPS AMONG STATISTICAL INDICES

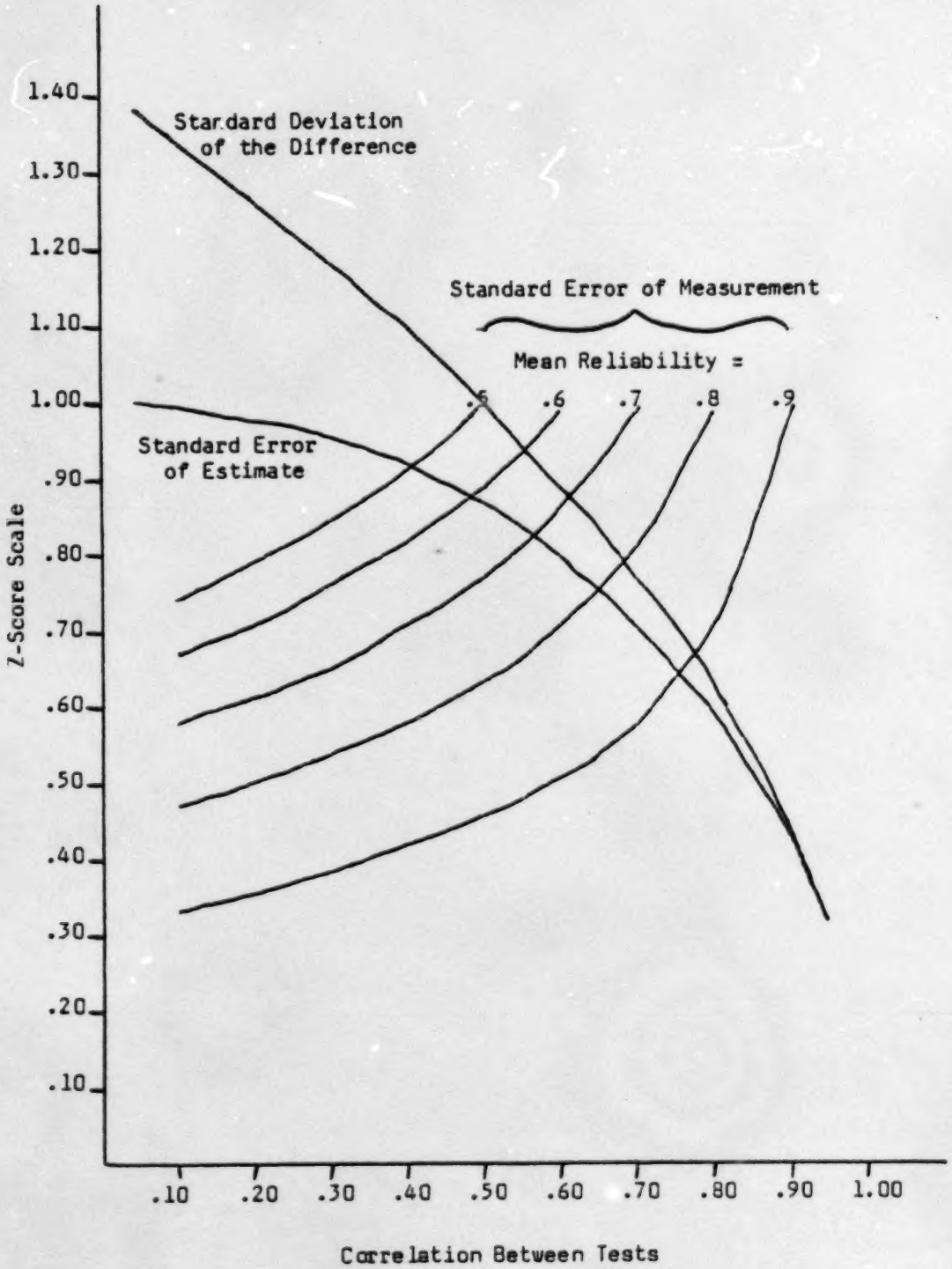


Figure 5 offers some clues to the psychometric limitations of any severe discrepancy model. The following points merit emphasis:

1. The correlation between the tests exerts strong effects on both the deviation of difference scores and the standard errors.
2. For tests whose correlation approaches their reliability, the standard error of measurement exceeds the standard deviation of the difference.
3. The standard error of estimate is smaller than the standard deviation of the difference, but for higher correlations (above about .7) they become nearly the same.
4. The standard errors are always less than 1.0; the standard deviation of the difference exceeds 1.0 only when the correlation is less than .5.
5. The smallest standard errors of measurement occur for highly reliable, uncorrelated tests; the largest standard errors of estimate occur for uncorrelated tests.

Discussion

Because the standard is supposed to represent a "severe" discrepancy, it appears theoretically sound to relate it to a discrepancy being unusual rather than merely statistically significant. Given reliable measures which are not colinear, many students may have statistically significant differences between ability and achievement which are not unusual with respect to the norm group. On the other hand, given unreliable or colinear tests even unusually large differences may not be statistically significant.

Regression models, especially those which incorporate information beyond just the ability score (for example, sex or socioeconomic status), may be able to predict achievement very accurately. The standard error of estimate will then be relatively small. It will also be identical, on the average, to the standard deviation of the differences between actual and expected achievement.

The major conceptual difference between the standard deviation of the difference and the standard error of estimate is that the standard deviation of the difference indicates that the ability/achievement difference is unusual with respect to the norm group. The standard error of estimate indicates that the achievement is unusually low given the students relative status (ability, SES) within the norm group. When the regression equation is based on the single variable, ability, it extends the standard deviation of the difference method to correct for regression to the mean. As other variables are added to the equation, the regression method begins the process of excluding alternative causes.

Because of the more complex corrections, standards based on regression methods are somewhat more difficult to explain than those based on the standard deviation of the difference. The qualifying difference for a given pair of tests under standard deviation of the difference approach is a constant. Under a standard error of estimate approach the qualifying difference varies with the level of ability and other characteristics. In addition to the practical difficulty of explaining the way in which the standard was set, there is a practical difficulty in obtaining the data

necessary to set it. Even accurate estimates of the overall correlation between two tests are difficult to obtain. Data on which to base regression estimates beyond the one variable (ability) model would be even more difficult to procure.

The State of Iowa has adopted a regression-based methodology using data from a two-year study conducted within the state. They explain:

> The regressed standard score procedure yields a distribution of regressed achievement scores that are predicted from levels of intellectual functioning. For any given level of intellectual functioning, the regressed procedure yields an expected or predicted mean achievement score. By using the expected mean achievement score, a criterion level for defining severe discrepancy can be established based on the standard deviation of the predicted achievement distribution. The regressed standard score procedure accounts for measurement error since the procedure utilizes the correlation of the two measures which is a function of the reliability of each measure. (Iowa, page 40)

They also caution that their data must be viewed as unique to their student population and not generalizable to other populations. This suggests that the establishment of a regression-based system might well require a prior data collection effort (this might also be true for any method which utilizes the correlation between the pairs of tests). The degree to which such research results can be expected to be stable over time also remains an unknown.

Conclusions

By moving from intuition-based expectancy formulae and years-below-grade-level standards to a standard deviation of the difference approach, the State of California has probably improved both the consistency and the accuracy of its severe discrepancy procedures. Practitioners utilizing this method can improve their accuracy by utilizing highly reliable measures whose correlation is at least .1 (and preferably .2) below the reliability of the less reliable test. They should also be aware that only very large discrepancy scores can be detected with adequate statistical power. For this reason, it is essential to have multiple indicators for each student. These indicators should be similar only with respect to their focus on the students' discrepancy. In all other ways (observer or test administrator, setting, time period, test used) they should be as independent as possible.

Finally, there is a critical need for a good data base which relates individually-administered ability and achievement measures for a full range population. Such a data base would give insight into the distribution of normality; the degree to which ability scores provide a ceiling to achievement scores, and the relationship between ability and variance in achievement. The statistical methodology necessary to make consistent, accurate and appropriate identification of severe discrepancies is available. The data base to which the methodology should be applied is not.

References

Bond, G. & Tinker, M. Reading Difficulties: Their Diagnosis and Correction (2nd Ed.). New York: Appleton-Century-Crofts, 1967.

Elliott, M. Quantative Evaluation Procedures for Learning Disabilities. Journal of Learning Disabilities, 1981, 14 (2), 84-87.

Erickson, M. T. The Z-Score Discrepancy Method for Identifying Reading-Disabled Children. Journal of Learning Disabilities, 1975, 8, 308-312.

Iowa, State of. The Identification of Pupils With Learning Disabilities. State of Iowa, Department of Public Instruction, Special Education Division, 1981.

Kirk, R. Introductory Statistics. California: Brooks/Cole Publishing Company, 1978.

McNemar, Q. Psychological Satistics (4th Ed.). New York: John Wiley and Sons, Inc., 1969.

Shepard, L. An Evaluation of the Regression Discrepancy Method For Identifying Children With Learning Disabilities. The Journal of Special Education, 1980, 14 (1), 79-91.